

Tong Zhou

✉ zhou.tong1@northeastern.edu | ☎ (734)-358-1431

[Google Scholar](#) | [LinkedIn](#) | [Homepage](#)

EDUCATION

Northeastern University, Boston, MA, USA

Sep. 2021 – present

Ph.D. in Electrical & Computer Engineering

University of Michigan, Ann Arbor, MI, USA

Sep. 2019 – May 2021

M.S. in Electrical & Computer Engineering

GPA: 3.81/4.0

Xidian University, Xi'an, Shaanxi, China

Sep. 2015 – Jul. 2019

B.S. in Electrical Engineering

GPA: 3.80/4.0

RESEARCH INTERESTS

AI Security · Privacy · Generative AI · Transfer Learning · Cryptography

TECHNICAL SKILLS

Programming: Python, MATLAB, Julia, C/C++

Frameworks & Others: PyTorch, TensorFlow, Numpy, Pandas, Scikit-learn, OpenCV

RESEARCH EXPERIENCE

Research Assistant @ Xiaolin Xu's Lab

Sep. 2021 – present

Advisor: *Prof. Xiaolin Xu*

Northeastern University

Exploring innovative solutions that mitigate risks and vulnerabilities in machine learning models, ultimately contributing to the advancement of trustworthy and responsible AI.

SELECTED PROJECTS

1. Enhancing Regulatory Measures for Generative Models (ongoing)

- This project responds to the pressing demand for regulatory measures amid the growing misuse of advanced generative models. With a focus on identifying the origin of generated content, our proposed framework ensures reliable detection of watermarks, immune to forging attempts by malicious parties.

2. Restrict Unauthorized Model Transfers at the Architecture Level (ICLR'24)

- Introduced an architecture-level defense against unauthorized transfers, ensuring optimal performance on source tasks while degrading performance on unauthorized tasks, regardless of attacker data access.
- Developed a zero-cost proxy-based binary predictor to accelerate Neural Architecture Search (NAS), incorporating task characteristics for efficient architecture assessment and enabling cross-task search with rank-based fitness scoring.

3. On-device Model IP Protection Leveraging Trusted Execution Environment (ICML'23)

- Systematically defined the requirements for active DNN model protection.
- Established a model IP protection system by partitioning the victim model into an obfuscated model and confidential model secrets. The latter is safeguarded by a Trusted Execution Environment, ensuring authorized inference.
- Designed an optimization algorithm to derive the obfuscated model by altering only a fraction of the victim model's weights; Performed experiments across various models and datasets, showcasing its resilience against adaptive model extraction attacks. This effectively hinders attackers from acquiring high-performing models.

4. Defend against DNN Architectural Extraction Attacks (ICCAD'22 Best Paper Nomination)

- Developed the pioneering framework to safeguard DNNs against architecture extraction attacks, exclusively through algorithm-level modifications to achieve DNN architecture protection.
- Designed a defense framework using NAS to counter model architectural extraction attacks; Introduced and evaluated seven obfuscation strategies to maintain the inference accuracy of the target model.

- Implemented the proposed framework and conducted experiments on NAS benchmarks, showcasing its superior performance compared to the current state-of-the-art obfuscation methods.

Research Assistant @ Jiande Chen’s Lab

Nov. 2020 – Apr. 2021

Advisor: Prof. Jiande Chen

University of Michigan

Developed deep learning models for feature extraction from electrocardiogram data to detect food intake phases, aiming to assist in treating obesity and diabetes.

Research Assistant @ Laboratory of Integrated Brain Imaging

May 2020 – Oct. 2020

Advisor: Prof. Zhongming Liu

University of Michigan

Enhanced segmentation performance for Transmission Electron Microscopy (TEM) images by integrating a self-attention mechanism into the U-Net architecture.

SELECTED PUBLICATIONS (*indicates equal contribution)

- ◇ **ArchLock: Locking DNN Transferability at the Architecture Level with a Zero-Cost Binary Predictor**
Tong Zhou, Shaolei Ren, and Xiaolin Xu
The Twelfth International Conference on Learning Representations (ICLR), 2024.
- ◇ **MirrorNet: A TEE-Friendly Framework for Secure On-device DNN Inference**
Ziyu Liu, Yukui Luo, Shijin Duan, **Tong Zhou** and Xiaolin Xu
IEEE/ACM International Conference on Computer-Aided Design (ICCAD), 2023.
- ◇ **AutoReP: Automatic ReLU Replacement for Fast Private Network Inference**
Hongwu Peng*, Shaoyi Huang*, **Tong Zhou***, Yukui Luo, Xiaolin Xu, Caiwen Ding, *et al.*
International Conference on Computer Vision (ICCV), 2023.
- ◇ **NNSplitter: An Active Defense Solution to DNN Model via Automated Weight Obfuscation**
Tong Zhou, Yukui Luo, Shaolei Ren, Xiaolin Xu
International Conference on Machine Learning (ICML), 2023.
- ◇ **ObfuNAS: A Neural Architecture Search-based DNN Obfuscation Approach**
Tong Zhou, Shaolei Ren, Xiaolin Xu
IEEE/ACM International Conference On Computer Aided Design (ICCAD), 2022.
Best Paper Nomination

SELECTED AWARDS

ICML Travel Grant	2023
COE Outstanding Graduate Student Award, Northeastern University	2023
IEEE/ACM William J. McCalla ICCAD Best Paper Nomination	2022
COE Dean’s Fellowship Award, Northeastern University	2021
Outstanding Graduate Award, Xidian University	2019
First Prize Scholarship, Xidian University	2016 - 2018

PROFESSIONAL SERVICE

Volunteer: ICML 2023, New England Hardware Security Workshop 2023
Reviewer: IEEE Systems Journal